Research Report

Linking With Continuous Exponential Families: Single-Group Designs

Shelby J. Haberman

November 2008

ETS RR-08-61

Linking With Continuous Exponential Families: Single-Group Designs

Shelby J. Haberman
ETS, Princeton, NJ

November 2008

Abstract

Continuous exponential families are applied to linking forms via a single-group design. In this application, a distribution from the continuous bivariate exponential family is used that has selected moments that match those of the bivariate distribution of scores on the forms to be linked. The selected continuous bivariate distribution then yields continuous univariate marginal distributions for the two forms. These marginal distributions then provide distribution functions and quantile functions that may be employed in equating. Normal approximations are obtained for the sample distributions of the conversion functions.

Key words: Moments, information theory

Acknowledgments

I thank Tim Davey, Dan Eignor, and Yi-Hsuan Lee for helpful comments.

Application of continuous exponential families to linking has been considered for equivalent-groups designs (Haberman, 2008). In such an application, it suffices to consider univariate continuous exponential families. For single-group designs for linking forms, bivariate continuous exponential families may be used. In this approach, the member of the bivariate continuous exponential family used to link two forms is chosen to share selected moments with the joint distribution of the scores on two forms to be linked. Once the bivariate continuous distribution is selected, one then has a continuous marginal distribution that corresponds to each of the forms under study. These marginal distributions may then be used to construct conversion functions that link the forms. This procedure is readily implemented. For random sampling, the estimated conversion functions are easily computed, and, for each point at which a conversion is desired, the estimated conversion function has an approximate normal distribution with mean 0 and an asymptotic covariance matrix that is readily estimated. Results in this report are related to recent work on an alternative to kernel equating (Wang, 2008), although the models in this report are more general and numerical methods are somewhat different.

Section 1 describes the proposed continuous exponential families for bivariate distributions and defines conversions based on these families are defined. Different models are compared by use of an information criterion (Haberman, 2008). Section 2 develops estimates of parameters and conversion functions and considers the large-sample properties of estimated conversions in terms of consistency and asymptotic normality. Estimated asymptotic standard deviations are provided for the estimated conversions. Section 3 presents a Newton-Raphson algorithm for computation of parameter estimates and discussed numerical quadrature issues.

Section 4 illustrates results by use of a published example to which kernel equating has previously been applied (von Davier, Holland, & Thayer, 2004).

Conclusions and possible further developments are examined in section 5.

1 Bivariate Continuous Exponential Families

In the equating problem considered, a sample of n examinees receives two different forms. For Form j, where j is 1 or 2, possible scores are in the closed interval with finite lower bound c_j and finite upper bound $d_j > c_j$. For examinee i, $1 \leq i \leq n$, the score on Form j is X_{ij}. It is assumed that the pairs (X_{i1}, X_{i2}), $1 \leq i \leq n$, are mutually independent and identically distributed random vectors such that X_{ij} has distribution function F_j. No requirement is imposed that the scores be integers or rational numbers. Nonetheless, in typical applications, F_j is not continuous, so that equating forms based on observed scores and some equipercentile approach normally involves some approximation of the distribution function F_j by a continuous distribution function G_j, which is strictly increasing on some open interval $B(j)$ that contains both c_j and d_j. For each positive real $p < 1$, there exists a unique continuous and increasing quantile function R_j such that $G_j(R_j(p)) = p$. The linking function e_{12} for conversion of a score on Form 1 to a score on Form 2 is then $e_{12}(x) = R_2(G_1(x))$ for x in $B(1)$, while the linking function e_{21} for conversion of a score on Form 2 to a score on Form 1 is $e_{21}(x) = R_1(G_2(x))$ for x in $B(2)$. Both e_{12} and e_{21} are strictly increasing and continuous on their respective ranges, and e_{12} and e_{21} are inverses, so that $e_{12}(e_{21}(x)) = x$ for x in $B(2)$ and $e_{21}(e_{12}(x)) = x$ for x in B_1 (Haberman, 2008). If g_1 is continuous at x in $B(1)$ and g_2 is continuous at $e_{12}(x)$, then application of standard results from calculus shows that e_{12} has derivative $e'_{12}(x) = g_1(x)/g_2(e_{12}(x))$ at x. Similarly, if g_2 is continuous at x in $B(2)$ and g_1 is continuous at $e_{21}(x)$, then e_{21} has derivative $e'_{21}(x) = g_2(x)/g_1(e_{21}(x))$ at x.

One method to obtain a suitable pair of distribution functions G_1 and G_2 is to approximate the joint distribution of $\mathbf{X}_i = (X_{i1}, X_{i2})$ by use of a bivariate continuous exponential family. Let $B(j)$ be bounded for $1 \leq j \leq 2$. Let u_{kj} be a polynomial of degree k on the interval $B(j)$ for $k \geq 0$ and $1 \leq j \leq 2$. For a pair $\mathbf{k} = (k(1), k(2))$ of nonnegative integers, let $u_{\mathbf{k}}$ be the polynomial on the plane such that $u_{\mathbf{k}}(\mathbf{x}) = u_{k(1)1}(x_1)u_{k(2)2}(x_2)$ for real pairs $\mathbf{x} = (x_1, x_2)$. Let $\mu_{\mathbf{k}}$ be the expectation of $u_{\mathbf{k}}(\mathbf{X}_i)$, so that $\mu_{\mathbf{k}}$ is a linear combination of the bivariate moments $E(X_{i1}^{j(1)} X_{i2}^{j(2)})$ of $\mathbf{X}_i$ for integers $j(1) \leq k(1)$ and $j(2) \leq k(2)$). Consider a nonempty set K of r pairs of nonnegative integers $\mathbf{k} = (k(1), k(2))$ such that $k(1)$ or $k(2)$ is positive. Let $\boldsymbol{\mu}_K$ be the K-array of $\mu_{\mathbf{k}}$, $\mathbf{k}$ in K, and let $\mathbf{u}_K(\mathbf{x})$ be the K-array of $u_{\mathbf{k}}(\mathbf{x})$, $\mathbf{k}$ in K. If $\mathbf{y}_K$ is a real K-array of $y_{\mathbf{k}}$, $\mathbf{k}$ in K, and $\mathbf{z}_K$ is a real K-array of $z_{\mathbf{k}}$, $\mathbf{k}$ in K, then let $\mathbf{y}'_K\mathbf{z}_K$ be the summation $\sum_{\mathbf{k}\in K} y_{\mathbf{k}} z_{\mathbf{k}}$. Assume that, for any real K-array $\mathbf{y}_K$, the variance of $\mathbf{y}'_K\mathbf{u}_K(\mathbf{X}_i)$ is 0 only if $y_{\mathbf{k}} = 0$ for each $\mathbf{k}$ in K. Let $B = B(1) \times B(2)$ be the interval in the plane that consists of pairs $(b(1), b(2))$ such that $b(1)$ is in $B(1)$ and $b(2)$ is in $B(2)$. Then a unique continuous bivariate distribution with positive density on B has the exponential family density

$$g_K(\mathbf{x}) = \gamma_K(\boldsymbol{\theta}_K) \exp[\boldsymbol{\theta}'_K \mathbf{u}_K(\mathbf{x})],$$

$\mathbf{x}$ in B, for a unique K-array $\boldsymbol{\theta}_K$ with elements $\theta_{\mathbf{k}K}$, $\mathbf{k}$ in K, and a unique positive real $\gamma_K(\boldsymbol{\theta}_K)$

such that

$$\int_B u_{\mathbf{k}}(\mathbf{x}) g_K(\mathbf{x}) d\mathbf{x} = \mu_{\mathbf{k}}$$

for $\mathbf{k}$ in K and

$$\int_B g_K(\mathbf{x}) d\mathbf{x} = 1$$

(Gilula & Haberman, 2000). A random vector $\mathbf{Y}_K = (Y_{1K}, Y_{2K})$ then exists such that Y_{jK} is in $B(j)$ for $1 \leq j \leq 2$ and $\mathbf{Y}_K$ has density g_K. The moment equalities $E(u_{\mathbf{k}}(\mathbf{Y}_K)) = E(u_{\mathbf{k}}(\mathbf{X}_i))$ hold for $\mathbf{k}$ in K, so that $\mathbf{Y}_K$ has a distribution close to that of $\mathbf{X}$ in the sense that the expected log penalty function $I_K = E(-\log g_K(\mathbf{X}))$ is the smallest expected log penalty function $E(-\log g(\mathbf{X}))$ for all probability densities g on B such that

$$g(\mathbf{x}) = \gamma_K(\boldsymbol{\theta}_{K*}) \exp[\boldsymbol{\theta}'_{K*} \mathbf{u}_K(\mathbf{x})]$$

for some real K-array $\boldsymbol{\theta}_{K*}$, and $E(-\log g(\mathbf{X})) = I_K$ only if $\boldsymbol{\theta}_{K*} = \boldsymbol{\theta}_K$.

The moment equations expressed in terms of $u_{\mathbf{k}}$ can be interpreted in terms of conventional moments if the set K satisfies the hierarchy rule that $(k(1), k(2))$ is in K whenever $(h(1), h(2))$ is in K, $k(1) \leq h(1)$, $k(2) \leq h(2)$, $k(1)$ and $k(2)$ are nonnegative integers, and $k(1)$ or $k(2)$ is positive. The equations $E(u_{\mathbf{k}}(\mathbf{Y}_K)) = E(u_{\mathbf{k}}(\mathbf{X}))$ for $\mathbf{k}$ in K then hold if, and only if, $E(Y_{1K}^{k(1)} Y_{2K}^{k(2)}) = E(X_{i1}^{k(1)} X_{i2}^{k(2)})$ for all $\mathbf{k}$ in K.

For $1 \leq j \leq 2$, the distribution function G_{jK} of Y_{jK} is strictly increasing and continuously differentiable on $B(j)$. If $B(j, y)$, y in $B(j)$, consists of all pairs (y_1, y_2) such that y_1 is in $B(1)$, y_2 is in $B(2)$, and $y_j \leq y$, then

$$G_{jK}(y) = \int_{B(j,y)} g(\mathbf{x}) d\mathbf{x}.$$

The inverse R_{jK} defined by $G_{jk}(R_{jK}(p))$ for $0 < p < 1$ is also continuously differentiable and strictly increasing, so that the conversion functions $e_{12K} = R_{2K}(G_{1K})$ and $e_{21K} = R_{1K}(G_{2K})$ are also continuously differentiable and strictly increasing.

As in the case of univariate exponential families (Haberman, 2008), numerical work is simplified if computations employ the Legendre polynomials P_k for $k \geq 0$ (Abramowitz & Stegun, 1965, chaps. 8, 22). These polynomials are determined by the equations $P_0(x) = 1$, $P_1(x) = x$, and

$$P_{k+1}(x) = (k+1)^{-1}[(2k+1)x P_k(x) - k P_{k-1}(x)],$$

$k \geq 1$. If $\inf(B(j))$ is the infimum of $B(j)$ and $\sup(B(j))$ is the supremum of $B(j)$, then it is relatively efficient for numerical work to let $\beta_j = [\inf(B(j)) + \sup(B(j))]/2$ be the midpoint of $B(j)$, $\eta_j = [\sup(B(j)) - \inf(B(j))]/2$ be half the range of $B(j)$, and

$$u_{kj}(x) = P_k((x - \beta_j)/\eta_j).$$

In applications considered in this report, for integers $r(j) > 1$, $1 \leq j \leq 2$, the set K consists of the $r(1) + r(2) + 1$ elements $(k(1), 0)$, $1 \leq k(1) \leq r(1)$, $(0, k(2))$, $1 \leq k(2) \leq r(2)$, and $(1, 1)$, so that the hierarchy principle holds, Y_{jK} and X_{ij} have the same $r(j)$ initial moments for $1 \leq j \leq 2$ and Y_{1K} and Y_{2K} have the same correlation as X_{i1} and X_{i2}. Thus Y_{jK} and X_{ij} have the same mean and variance for each j. If $r(j) > 2$, then Y_{jK} and X_{ij} have the same skewness coefficient. If $r(j) > 3$, then Y_{jK} and X_{ij} have the same kurtosis coefficient. In the case of $r(1) = r(2) = 2$ and use of Legendre polynomials, if $\theta_{\mathbf{k}}$ is negative for $\mathbf{k}$ equal to $(2, 0)$ or $(0, 2)$ and $\theta^2_{(1,1)}$ is less than $36\theta^2_{(2,0)}\theta^2_{(0,2)}$, then $\mathbf{Y}_K$ corresponds to a bivariate normal random variable $\mathbf{Z} = (Z_1, Z_2)$. The distribution of $\mathbf{Y}_K$ is the same as the conditional distribution of $\mathbf{Z}$ conditional on Z_j being in $B(j)$ for $1 \leq j \leq 2$. Comparison of the density formula for the bivariate normal with the density formula for the corresponding continuous exponential family shows that the unconditional variance of Z_1 satisfies

$$\sigma^2(Z_1) = -\frac{\eta_1^2}{6\theta_{(2,0)} - \theta^2_{(1,1)}/(6\theta_{(0,2)})},$$

the unconditional variance of Z_2 is

$$\sigma^2(Z_2) = -\frac{\eta_2^2}{6\theta_{(0,2)} - \theta^2_{(1,1)}/(6\theta_{(2,0)})},$$

the unconditional correlation of Z_1 and Z_2 is

$$\rho(Z_1, Z_2) = \theta_{(1,1)}/[6(\theta_{(2,0)}\theta_{(0,2)})^{1/2}],$$

the unconditional mean of Z_1 is

$$E(Z_1) = \beta_1 - \frac{\eta_1[\theta_{(1,0)} - \theta_{(0,1)}\theta_{(1,1)}/(6\theta_{(0,2)})]}{6\theta_{(2,0)} - \theta^2_{(1,1)}/(6\theta_{(0,2)})},$$

and the unconditional mean of Z_2 is

$$E(Z_2) = \beta_2 - \frac{\eta_2[\theta_{(0,1)} - \theta_{(1,0)}\theta_{(1,1)}/(6\theta_{(2,0)})]}{6\theta_{(0,2)} - \theta^2_{(1,1)}/(6\theta_{(2,0)})}.$$

One alternative choice of K (Wang, 2008) has K contain all pairs $(k(1), k(2))$ of nonnegative integers such that $k(1)$ or $k(2)$ is positive, $k(1) \leq r(1)$, and $k(2) \leq r(2)$. If $r(1) > 1$ and $r(2) > 2$, then $E(Y_{1K}^{k(1)} Y_{2K}^{k(2)}) = E(X_{i1}^{k(1)} X_{i2}^{k(2)})$ for $1 \leq k(1) \leq r(1)$ and $1 \leq k(2) \leq r(2)$.

2 Estimation of Parameters

The parameter $\boldsymbol{\theta}_K$, the information criterion I_K, the distribution functions G_{jK}, and the conversion functions e_{12K} and e_{21K} are readily estimated (Gilula & Haberman, 2000; Haberman, 2008). For $\mathbf{k}$ in K, let $m_{\mathbf{k}}$ be the sample mean $n^{-1} \sum_{i=1}^{n} u_{\mathbf{k}}(\mathbf{X}_i)$, and let $\mathbf{m}_K$ be the K-array with elements $m_{\mathbf{k}}$, $\mathbf{k}$ in K . If the covariance matrix of $\mathbf{m}_K$ is positive definite, then $\boldsymbol{\theta}_K$ is estimated by the unique K-array $\hat{\boldsymbol{\theta}}_K$ such that

$$\int_B \mathbf{u}_K(\mathbf{x}) \hat{g}_K(\mathbf{x}) d\mathbf{x} = \mathbf{m}_K,$$

$$\int_B \hat{g}_K(\mathbf{x}) d\mathbf{x} = 1,$$

and

$$\hat{g}_K(\mathbf{x}) = \gamma(\hat{\boldsymbol{\theta}}_K) \exp[\hat{\boldsymbol{\theta}}_K' \mathbf{u}_K(\mathbf{x})]$$

for $\mathbf{x}$ in B.

As the sample size n approaches ∞, $\hat{\boldsymbol{\theta}}_K$ converges to $\boldsymbol{\theta}_K$ with probability 1, and $n^{1/2}(\hat{\boldsymbol{\theta}}_K - \boldsymbol{\theta}_K)$ converges in distribution to a multivariate normal random variable with 0 mean and with covariance matrix $\mathbf{A}_K = \mathbf{C}_K^{-1} \mathbf{D}_K \mathbf{C}_K^{-1}$ (Gilula & Haberman, 2000). Here $\mathbf{D}_K$ is the covariance matrix of $\mathbf{u}_K(\mathbf{X})$ and $\mathbf{C}_K$ is the covariance matrix of the K-array $\mathbf{u}_K(\mathbf{Y}_K)$. Thus

$$\mathbf{C}_K = \int_B [\mathbf{u}_K(\mathbf{x}) - \boldsymbol{\mu}_K][\mathbf{u}_K(\mathbf{x}) - \boldsymbol{\mu}_K]' g_K(\mathbf{x}) d\mathbf{x}.$$

The estimate of $\mathbf{C}_K$ is

$$\hat{\mathbf{C}}_K = \int_B [\mathbf{u}_K(\mathbf{x}) - \mathbf{m}_K][\mathbf{u}_K(\mathbf{x}) - \mathbf{m}_K]' \hat{g}_K(\mathbf{x}) d\mathbf{x}.$$

The estimate of $\mathbf{D}_K$ is

$$\hat{\mathbf{D}}_K = (n-1)^{-1} \sum_{i=1}^{n} [\mathbf{u}_K(\mathbf{X}_i) - \mathbf{m}_K][\mathbf{u}_K(\mathbf{X}_i) - \mathbf{m}_K]'.$$

Thus $\mathbf{A}_K$ has estimate

$$\hat{\mathbf{A}}_K = \hat{\mathbf{C}}_K^{-1} \hat{\mathbf{D}}_K \hat{\mathbf{C}}_K^{-1}.$$

For any nonzero constant K-array $\mathbf{z}_K$, the estimated asymptotic standard deviation (EASD) of $\mathbf{z}_K'\hat{\boldsymbol{\theta}}_K$ is

$$\hat{\sigma}(\mathbf{z}_K'\hat{\boldsymbol{\theta}}_K) = n^{-1/2}(\mathbf{z}_K'\hat{\mathbf{A}}_K\mathbf{z}_K)^{1/2},$$

so that $(\mathbf{z}_K'\hat{\boldsymbol{\theta}}_K - \mathbf{z}_K'\boldsymbol{\theta}_K)/\hat{\sigma}(\mathbf{z}_K'\hat{\boldsymbol{\theta}}_K)$ converges in distribution to a standard normal random variable.

The minimum expected penalty I_K may be estimated by

$$\hat{I}_K = -\log\gamma_K(\hat{\boldsymbol{\theta}}_K) - \hat{\boldsymbol{\theta}}_K'\mathbf{m}_K.$$

As the sample size n increases, $\hat{I}_K$ converges to I_K with probability 1 and $n^{1/2}(\hat{I}_K - I_K)$ converges in distribution to a normal random variable with mean 0 and variance

$$\sigma^2(-\log g_K(\mathbf{X})) = \mu_K'\mathbf{A}_K\mu_K.$$

The EASD of $\hat{I}_K$ is then

$$\hat{\sigma}(\hat{I}_K) = n^{-1/2}(\mathbf{m}_K'\hat{\mathbf{A}}_K\mathbf{m}_K)^{1/2}.$$

To verify this claim, let H_K be the function defined by

$$H_K(\boldsymbol{\theta}) = -n^{-1}\sum_{i=1}^{n}\log g_K(\boldsymbol{\theta}),$$

and let $H_{K0} = K_K(\boldsymbol{\theta}_K)$. Apply the central limit theorem to H_{K0} to show that the scaled difference $n^{1/2}(H_{K0} - I_K)$ converges in distribution to a normal random variable with mean 0 and variance $\sigma^2(-\log g_K(\mathbf{X}))$. Differentiation and Taylor's theorem show that the difference

$$2(H_{K0} - \hat{I}_K) = (\hat{\boldsymbol{\theta}}_K - \boldsymbol{\theta}_K)'\mathbf{C}_K^*(\hat{\boldsymbol{\theta}}_K - \boldsymbol{\theta}_K)$$

for $\mathbf{C}_K^*$ the covariance matrix of the K-array $\mathbf{u}_K(\mathbf{Y}_K^*)$ for some random vector $\mathbf{Y}_K^*$ with density $\gamma_K(\boldsymbol{\theta}_K^*)\exp[(\boldsymbol{\theta}_K^*)'\mathbf{u}_K(\mathbf{x})]$ for some $\boldsymbol{\theta}_K^*$ on the line segment between $\boldsymbol{\theta}_K$ and $\hat{\boldsymbol{\theta}}_K$. It follows that $n^{1/2}(H_{K0} - \hat{I}_K)$ converges in probability to 0 and $n^{1/2}(\hat{I}_K - I_K)$ converges in distribution as claimed.

For $1 \le j \le 2$, the distribution function G_{jK} has estimate $\hat{G}_{jK}$ defined by

$$\hat{G}_{jK}(y) = \int_{B(j,y)}\hat{g}_K(\mathbf{x})d\mathbf{x}$$

for y in $B(j)$, and the quantile function R_{jK} has estimate $\hat{R}_{jK}$ defined by

$$\hat{G}_{jK}(\hat{R}_{jK}(p)) = p$$

for $0 < p < 1$. For a continuously differentiable function f_{Kjy} on the set of K arrays, $\hat{G}_{jk}(y) = f_{Kjy}(\hat{\boldsymbol{\theta}}_K)$, and f_{Kjy} has gradient

$$\mathbf{T}_{jK}(y) = \int_{B(j,y)} [\mathbf{u}_K(\mathbf{x}) - \boldsymbol{\mu}_K] g_K(\mathbf{x}) d\mathbf{x}$$

at $\boldsymbol{\theta}_K$. The delta method (Rao, 1973, p. 388) and the fact that $\hat{G}_{jK}$ is a continuous distribution function may be employed to demonstrate that, as the sample size n approaches ∞, $\hat{G}_{jK}(y)$ converges to $G_{jK}(y)$ with probability 1 for y in $B(j)$, so that $|\hat{G}_{jK} - G_{jK}|$, the supremum of $|\hat{G}_{jK}(y) - G_{jK}(y)|$ for y in $B(j)$, converges to 0 with probability 1. In addition, $[\hat{G}_{Kj}(y) - G_{jK}(y)]/\sigma(\hat{G}_{jK}(y))$ converges in distribution to a normal random variable with mean 0 and variance 1 if

$$\sigma(\hat{G}_{jK}(y)) = n^{-1/2}\{[\mathbf{T}_{jK}(y)]'\mathbf{A}_K\mathbf{T}_{jK}(y)\}^{1/2}.$$

Similarly, $\hat{R}_{jK}(p)$ converges to $R_{jK}(p)$ with probability 1, and $[\hat{R}_{jK}(p) - R_{jK}(p)]/\sigma(\hat{R}_{jK}(p))$ converges in distribution to a normal random variable with mean 0 and variance 1 if

$$\sigma(\hat{R}_{jK}(p)) = [g_{jK}(R_{jK}(p))]^{-1}\sigma(\hat{G}_{jK}(R_{jK}(p)))$$

and $g_{jK}(y)$ is the marginal density corresponding to G_{jK}. Thus $g_{1K}(y)$ is the integral of $g_K((y, x_2))$ over x_2 in $B(2)$, and $g_{2K}(y)$ is the integral of $g_K((x_1, y))$ over x_1 in $B(1)$. Estimated asymptotic standard deviations may be derived by use of obvious substitutions of estimated parameters for actual parameters. Thus

$$\hat{\sigma}(\hat{G}_{jK}(y)) = n^{-1/2}\{[\hat{\mathbf{T}}_{jK}(y)]'\hat{\mathbf{A}}_K\hat{\mathbf{T}}_{jK}(y)\}^{1/2},$$

where

$$\hat{\mathbf{T}}_{jK}(y) = \int_{B(j,y)} [\mathbf{u}_K(\mathbf{x}) - \mathbf{m}_K] \hat{g}_K(\mathbf{x}) d\mathbf{x},$$

$$\hat{\sigma}(\hat{R}_{jK}(p)) = [\hat{g}_{jK}(\hat{R}_{jK}(p))]^{-1}\hat{\sigma}(\hat{G}_{jK}(\hat{R}_{jK}(p))),$$

and $\hat{g}_{jK}(y)$ is the marginal density corresponding to $\hat{G}_{jK}$.

The estimate $\hat{e}_{12K}$ of the conversion function $e_{12K}(y)$ from Form 1 to Form 2 satisfies $\hat{e}_{12K}(y) = \hat{R}_{2K}(\hat{G}_{1K}(y))$ for y in $B(1)$, and the corresponding estimate $\hat{e}_{21K}$ of e_{21K} satisfies $\hat{e}_{21K}(y) = \hat{R}_{1K}(\hat{G}_{2K}(y))$ for y in $B(2)$ As the sample size n become large, $\hat{e}_{12K}(y)$ converges with probability 1 to $e_{12K}(y)$ for y in $B(1)$, and $\hat{e}_{21K}(y)$ converges with probability 1 to $e_{21K}(y)$ for

y in $B(2)$. In addition, $[\hat{e}_{12K}(y) - e_{12K}(y)]/\sigma(\hat{e}_{12K}(y))$ converges in distribution to a standard normal random variable if

$$\sigma(\hat{e}_{12}(y)) = n^{-1/2}\{[\mathbf{T}_{1K}(y) - \mathbf{T}_{2K}(e_{12}(y))]'\mathbf{A}_K[\mathbf{T}_{1K}(y) - \mathbf{T}_{2K}(e_{12}(y))]\}^{1/2},$$

In like manner, $[\hat{e}_{21K}(y) - e_{21K}(y)]/\sigma(\hat{e}_{21K}(y))$ converges in distribution to a standard normal random variable if

$$\sigma(\hat{e}_{21}(y)) = n^{-1/2}\{[\mathbf{T}_{2K}(y) - \mathbf{T}_{1K}(e_{21}(y))]'\mathbf{A}_K[\mathbf{T}_{2K}(y) - \mathbf{T}_{1K}(e_{21}(y))]\}^{1/2},$$

The EASD of $\hat{e}_{12}(y)$ is

$$\hat{\sigma}(\hat{e}_{12}(y)) = n^{-1/2}\{[\hat{\mathbf{T}}_{1K}(y) - \hat{\mathbf{T}}_{2K}(\hat{e}_{12}(y))]'\hat{\mathbf{A}}_K[\hat{\mathbf{T}}_{1K}(y) - \hat{\mathbf{T}}_{2K}(\hat{e}_{12}(y))]\}^{1/2},$$

The EASD of $\hat{e}_{21}(y)$ is

$$\hat{\sigma}(\hat{e}_{21}(y)) = n^{-1/2}\{[\hat{\mathbf{T}}_{2K}(y) - \hat{\mathbf{T}}_{1K}(\hat{e}_{21}(y))]'\hat{\mathbf{A}}_K[\hat{\mathbf{T}}_{2K}(y) - \hat{\mathbf{T}}_{1K}(\hat{e}_{21}(y))]\}^{1/2}.$$

3 Computational Issues

Given a starting value $\boldsymbol{\theta}_{K0}$, the Newton-Raphson algorithm may be employed to compute $\hat{\boldsymbol{\theta}}_K$. Computation of $\hat{\boldsymbol{\theta}}_K$ corresponds to minimization of H_K. At step $t \geq 0$, a new approximation $\boldsymbol{\theta}_{K(t+1)}$ of $\hat{\boldsymbol{\theta}}_K$ is found by the equation

$$\boldsymbol{\theta}_{K(t+1)} = \boldsymbol{\theta}_{Kt} + \mathbf{C}_{Kt}^{-1}[\mathbf{m}_K - \boldsymbol{\mu}_{Kt}].$$

Here

$$g_{Kt}(\mathbf{x}) = \gamma_K(\boldsymbol{\theta}_{Kt})\exp[\boldsymbol{\theta}'_{Kt}\mathbf{u}_K(\mathbf{x})],$$

$\mathbf{x}$ in B, $\gamma_K(\boldsymbol{\theta}_{Kt})$ is defined so that

$$\int_B g_{Kt}(\mathbf{x})d\mathbf{x} = 1,$$

$$\boldsymbol{\mu}_{Kt} = \int_B \mathbf{u}_K(\mathbf{x})g_{Kt}(\mathbf{x})d\mathbf{x},$$

$\mathbf{m}_K - \boldsymbol{\mu}_{Kt}$ is the gradient of H_K at $\boldsymbol{\theta}_{Kt}$, and

$$\mathbf{C}_{Kt} = \int_B [\mathbf{u}_K(\mathbf{x}) - \boldsymbol{\mu}_{Kt}][\mathbf{u}_K(\mathbf{x}) - \boldsymbol{\mu}_{Kt}]'g_{Kt}(\mathbf{x})d\mathbf{x}$$

is the Hessian matrix of H_K at $\boldsymbol{\theta}_{Kt}$.

The Newton-Raphson algorithm is also employed to evaluate quantile functions (Haberman, 1996, pp. 426–427). The algorithm considers solution of the equation

$$\log[\hat{G}_{jK}(\hat{R}_{jK}(p))/p] = 0$$

for a given p in the interval $(0, 1)$. Given an initial approximation $R_{jk0}(p)$ to $\hat{R}_{jK}(p)$ and the fact that $\log(\hat{G}_{jK})$ has derivative $\hat{g}_{jK}/\hat{G}_{jK}$, the Newton-Raphson algorithm produces approximations $R_{jKt}(p)$ to $\hat{R}_{jK}(p)$ such that

$$R_{jK(t+1)}(p) = R_{jKt}(p) - \hat{G}_{jK}(R_{jKt}(p)) \log[\hat{G}_{jK}(R_{jK}(p))/p]/\hat{g}_{jK}(R_{jKt}(p)).$$

The Legendre polynomials used in typical cases to define $\mathbf{u}_K$ also form the basis for the Gaussian quadrature used for evaluation of all integrals on B, $B(1)$, and $B(2)$ (Abramowitz & Stegun, 1965, p. 887). In the case of $\hat{G}_{jK}(y)$, the limits of integration of the numerator related to Form j are from $\inf(B(j))$ to y rather than from $\inf(B(j))$ to $\sup(B(j))$, so that the scaled Legendre polynomials are modified accordingly for Gaussian quadrature. In this paper, calculations use 20-point Gaussian quadrature for each dimension. The larger number of points than used in previous work on equivalent groups (Haberman, 2008) reflects the numerical challenge associated with form scores that are typically highly correlated.

The rectangle rule of integration has been employed in the literature for the case with $B(j) = (c_j - 0.5, d_j + 0.5)$ (Wang, 2008), with c_j and d_j integers, and with the integers from c_j to d_j used as equally weighted quadrature points. Because the distance between quadrature points is not small, this integration approach yields errors that are not negligible. For expectations $E(Y_{1K}^{k(1)} Y_{2K}^{k(2)})$, approximation errors are of order $(d_1 - c_1)^{-2} + (d_2 - c_2)^{-2}$, so that they can be expected to be relatively minor for very long tests and somewhat more obvious for quite short tests. Unfortunately, the proper conditions for use of Sheppard's corrections (Kolassa & McCullagh, 1990) do not apply, so that the most precise approach to evaluation of errors is unavailable. It should be emphasized that, in practice, use of Legendre polynomials results in both higher accuracy and less computation than does use of the rectangle rule.

4 Example

Table 8.2 of von Davier et al. (2004) provides an example of a single-group design with $c_j = 0$ and $d_j = 20$ for $1 \leq j \leq 2$. To illustrate results, let $B(1) = B(2) = (-0.5, 20.5)$. The u_{kj}

are defined based on Legendre polynomials, and three sets K are considered. These sets are

$$K(2) = \{(1,0),(2,0),(0,1),(0,2),(1,1)\},$$

$$K(3) = \{(1,0),(2,0),(3,0),(0,1),(0,2),(0,3),(1,1)\},$$

and

$$K(4) = \{(1,0),(2,0),(3,0),(4,0),(0,1),(0,2),(0,3),(0,4),(1,1)\}.$$

Results for parameters are summarized in Table 1. Results in terms of estimated expected log penalties are summarized in Table 2. These results suggest that gains beyond the quadratic case $K(2)$ are quite small, although $K(4)$ differs from $K(3)$ more than $K(3)$ differs from $K(2)$.

Table 1
Model Parameters

	$K(2)$		$K(3)$		$K(4)$	
k	Est.	EASD	Est.	EASD	Est.	EASD
$(1,0)$	0.902	0.118	1.012	0.140	1.056	0.153
$(2,0)$	-6.174	0.226	-6.226	0.231	-6.381	0.262
$(3,0)$			0.174	0.124	0.234	0.147
$(4,0)$					-0.192	0.152
$(0,1)$	-0.463	0.126	-0.543	0.138	-0.477	0.179
$(0,2)$	-7.152	0.266	-7.141	0.265	-7.931	0.307
$(0,3)$			-0.117	0.117	-0.034	0.167
$(0,4)$					-0.836	0.164
$(1,1)$	15.619	0.663	15.624	0.664	15.520	0.661

Note. EASD = estimated asymptotic standard deviation, est. = estimate.

Table 2
Estimated Expected Log Penalties

K	Estimate	EASD
$K(2)$	4.969	0.022
$K(3)$	4.968	0.022
$K(4)$	4.960	0.022

Note. EASD = estimated asymptotic standard deviation, est. = estimate.

Not surprisingly, the three choices of K lead to rather similar conversion functions. Consider Table 3 for the case of conversion of Form 1 to Form 2. A bit more variability in results exists for very high or very low values, although estimated asymptotic standard deviations are more variable than are estimated conversions. Note that results are also similar to those for kernel equating (von Davier et al., 2004, chap. 8) shown in Table 4. These results employ a log-linear model for the joint distribution of the scores that is comparable to the model defined by $K(3)$ for a continuous exponential family. The log-linear fit preserves the initial three marginal moments for each score distribution as well as the covariance of the two scores. As a consequence, the marginal distributions produced by the kernel method have the same means and variances as do the corresponding distributions of X_{i1} and X_{i2}, but the kernel methods yields the distribution of a continuous random variable for the first form with a skewness coefficient that is 0.987 times the original skewness coefficient for X_{i1} and a distribution of a continuous random variable for the second form that with a skewness coefficient that is 0.983 times the original skewness coefficient for X_{i2}.

5 Conclusions

As in the case of equivalent groups (Haberman, 2008), linking forms in a single-group design by continuous exponential families appears similar in result to linking the same forms via kernel equating. Continuous exponential families offer some possible gains. Unlike kernel equating, bandwidths are not required, so that fewer specifications are required. In kernel equating, log-linear smoothing and production of continuous distribution functions require distinct steps. In the case of continuous exponential families, a model is fit that immediately results in continuous distribution functions.

Although this gain is not apparent in the example, a possible gain from continuous exponential families is that application to assessments with unevenly spaced scores or very large numbers of possible scores is completely straightforward. Thus direct conversion from a raw score on one form to an unrounded scale score on a second form involves no difficulties. In addition, in tests with formula scoring, no need exists to round raw scores to integers during equating.

The single-group design provides the basis for more complex linking designs with anchor tests (von Davier et al., 2004, chap. 9), so that results of this report are readily applied to a very wide variety of linking problems.

Table 3

Comparison of Conversions From Form 1 to Form 2

	$K(2)$		$K(3)$		$K(4)$	
Value	Est.	EASD	Est.	EASD	Est.	EASD
0	0.111	0.077	-0.040	0.113	0.404	0.262
1	1.168	0.128	0.927	0.204	1.404	0.264
2	2.144	0.135	1.917	0.208	2.269	0.221
3	3.091	0.130	2.910	0.182	3.121	0.176
4	4.028	0.120	3.899	0.151	3.987	0.140
5	4.959	0.108	4.881	0.122	4.874	0.117
6	5.889	0.097	5.854	0.101	5.785	0.105
7	6.819	0.086	6.819	0.087	6.721	0.100
8	7.748	0.076	7.775	0.080	7.679	0.095
9	8.677	0.069	8.722	0.078	8.653	0.090
10	9.606	0.065	9.661	0.078	9.634	0.086
11	10.536	0.064	10.591	0.077	10.611	0.085
12	11.465	0.066	11.512	0.077	11.574	0.088
13	12.394	0.073	12.425	0.077	12.514	0.092
14	13.324	0.081	13.331	0.081	13.427	0.094
15	14.256	0.091	14.231	0.090	14.310	0.096
16	15.193	0.102	15.128	0.105	15.166	0.101
17	16.141	0.113	16.033	0.126	16.003	0.116
18	17.119	0.121	16.967	0.149	16.838	0.142
19	18.173	0.123	17.985	0.167	17.716	0.174
20	19.495	0.094	19.304	0.150	18.842	0.198

Note. EASD = estimated asymptotic standard deviation, est. = estimate.

Table 4
Conversions From Form 1 to Form 2 by Kernel Equating

Value	Estimate	EASD
0	-0.002	0.162
1	0.999	0.221
2	1.981	0.221
3	2.956	0.193
4	3.926	0.159
5	4.890	0.128
6	5.850	0.104
7	6.805	0.089
8	7.756	0.080
9	8.702	0.078
10	9.643	0.077
11	10.580	0.077
12	11.512	0.077
13	12.439	0.078
14	13.362	0.083
15	14.283	0.095
16	15.206	0.115
17	16.140	0.140
18	17.105	0.167
19	18.155	0.185
20	19.411	0.158

Note. EASD = estimated asymptotic standard deviation, est. = estimate.

References

Abramowitz, M., & Stegun, I. A. (1965). *Handbook of mathematical functions.* New York: Dover.

Gilula, Z., & Haberman, S. J. (2000). Density approximation by summary statistics: An information-theoretic approach. *Scandinavian Journal of Statistics*, *27*, 521–534.

Haberman, S. J. (1996). *Advanced statistics: Vol. 1. Description of populations.* New York: Springer-Verlag.

Haberman, S. J. (2008). *Continuous exponential families: An equating tool* (Research Rep. No. RR-08-05). Princeton, NJ: ETS.

Kolassa, J. E., & McCullagh, P. (1990). Edgeworth series for lattice distributions. *The Annals of Statistics*, *18*, 981–985.

Rao, C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). New York: John Wiley.

von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). *The kernel method of test equating.* New York: Springer.

Wang, T. (2008). The continuized log-linear method: An alternative to the kernel method of continuization in test equation. *Applied Psychological Measurement*, *33*(7), 527–542.